**SVEUČILIŠTE JOSIPA JURJA STROSSMAYERA U OSIJEKU**

**FAKULTET ELEKTROTEHNIKE, RAČUNARSTVA I INFORMACIJSKIH TEHNOLOGIJA**

**Procjena cijene rabljenih mobilnih uređaja**

**Seminarski rad**

**Raspoznavanje uzoraka i strojno učenje**

**Martin Marenjak**

**Osijek, 2024.**

Sadržaj

[1. UVOD 2](#_Toc168501505)

[2. NADZIRANO UČENJE 3](#_Toc168501506)

[2.1. Regresija 4](#_Toc168501507)

[3. ANALIZA PODATAKA 6](#_Toc168501508)

[3.1. Eksplorativna analiza podataka 9](#_Toc168501509)

[4. Priprema podataka 13](#_Toc168501510)

[4.1. Podatci koji nedostaju 13](#_Toc168501511)

[4.2. Inženjering značajki 15](#_Toc168501512)

[4.3. Priprema podataka za modeliranje 16](#_Toc168501513)

[5. Evaluacija rješenja 17](#_Toc168501514)

[5.1. Linearna regresija 17](#_Toc168501515)

[5.2. Random Forest Regressor 18](#_Toc168501516)

[5.3. Stablo odlučivanja 20](#_Toc168501517)

[5.4. K-najbližih susjeda 21](#_Toc168501518)

[6. Zaključak 23](#_Toc168501519)

# UVOD

U današnjem digitalnom dobu, mobilni uređaji igraju ključnu ulogu u svakodnevnom životu, što je dovelo do sve veće potražnje i raznolikosti na tržištu. Procjena cijene mobilnih uređaja postaje sve složenija s obzirom na brojne faktore koji utječu na njihovu vrijednost. U sklopu kolegija Raspoznavanje uzoraka i strojno učenje, cilj ovog seminarskog rada je razviti model koji će predviđati cijenu rabljenih mobilnih uređaja temeljem različitih karakteristika i značajki uređaja.

Za potrebe ovog projekta koristit će se skup podataka koji sadrži informacije o rabljenim mobilnim uređajima, uključujući model uređaja, procesor, RAM, kameru, bateriju, godinu proizvodnje, broj dana korištenja i druge relevantne značajke. Eksplorativna analiza podataka obuhvatit će statističke metrike i vizualizacije kako bi se bolje razumjela distribucija i međusobna povezanost značajki unutar skupa podataka.

U okviru ovog rada, bit će primijenjene metode strojnog učenja poput linearne regresije i stabla odlučivanja kako bi se razvili modeli za predikciju cijene mobilnih uređaja. Evaluacija razvijenih modela izvršit će se usporedbom različitih metrika performansi, čime će se omogućiti identifikacija najpreciznijeg pristupa za ovaj zadatak.

Ovaj projekt ne samo da će pružiti uvid u trenutne metode procjene cijene rabljenih mobilnih uređaja, već će također demonstrirati praktičnu primjenu tehnika strojnog učenja u analizi i predikciji vrijednosti proizvoda na tržištu tehnologije.

# NADZIRANO UČENJE

Nadzirano učenje je grana strojnog učenja koja se fokusira na učenje iz unaprijed označenih podataka. U ovoj vrsti učenja, model se trenira na skupu podataka koji sadrži ulazne podatke (značajke) i odgovarajuće izlazne vrijednosti (oznake). Cilj je naučiti mapiranje između ulaznih i izlaznih podataka kako bi model mogao precizno predvidjeti izlazne vrijednosti za nove, nepoznate ulazne podatke.

Proces nadziranog učenja započinje prikupljanjem skupa podataka koji se dijeli na skup za treniranje i skup za testiranje. Skup za treniranje koristi se za izgradnju modela, dok se skup za testiranje koristi za procjenu njegove točnosti i sposobnosti generalizacije. Algoritmi nadziranog učenja prilagođavaju svoje parametre tijekom procesa treniranja tako da minimiziraju razliku između predviđenih i stvarnih vrijednosti u skupu za treniranje.

Postoje dvije glavne kategorije nadziranog učenja: regresija i klasifikacija. Regresija se koristi kada je izlazna varijabla kontinuirana, kao što je predviđanje cijene mobilnog uređaja. Klasifikacija se koristi kada je izlazna varijabla diskretna, kao što je klasificiranje vrste cvijeta na temelju njegovih značajki.

Popularni algoritmi nadziranog učenja uključuju linearnu regresiju, logističku regresiju, stabla odlučivanja, podržavajuće vektorske strojeve (SVM) i neuronske mreže. Svaki od ovih algoritama ima svoje prednosti i nedostatke te je izbor algoritma često ovisan o prirodi problema i skupa podataka.

Evaluacija modela nadziranog učenja ključna je za razumijevanje njegove učinkovitosti. Uobičajene metrike za regresijske probleme uključuju srednju apsolutnu pogrešku (MAE), srednju kvadratnu pogrešku (MSE) i koeficijent determinacije (R²). Za klasifikacijske probleme koriste se metrike kao što su točnost, preciznost, odziv i F1 mjera.

Nadzirano učenje je temelj mnogih praktičnih aplikacija strojnog učenja, uključujući prepoznavanje slike, obradu prirodnog jezika i prediktivnu analitiku. Njegova sposobnost da iz velikih količina podataka izvuče vrijedne uvide čini ga neizostavnim alatom u modernoj analizi podataka.

## Regresija

Regresija je ključna metoda nadziranog učenja koja se koristi za predviđanje kontinuiranih vrijednosti. U nadziranom učenju, model se trenira na skupu podataka koji sadrži ulazne značajke i pripadajuće izlazne vrijednosti, s ciljem učenja mapiranja između tih ulaza i izlaza. Regresija je fokusirana na kvantitativne prognoze, gdje izlazne vrijednosti nisu diskretne kategorije, već kontinuirani brojevi.

Jedan od osnovnih i najčešće korištenih regresijskih modela je linearna regresija. Linearna regresija pokušava pronaći najbolju linearnu vezu između ulaznih značajki i izlazne varijable, minimizirajući razliku između stvarnih i predviđenih vrijednosti. Ovo omogućava modelu da precizno predvidi izlazne vrijednosti za nove ulazne podatke.

Osim linearne regresije, drvo odlučivanja je još jedan važan model za regresiju. Drvo odlučivanja koristi strukturu sličnu stablu za donošenje odluka, gdje se podaci dijele prema određenim značajkama kako bi se došlo do predikcije. Ova metoda je vrlo intuitivna i omogućava vizualizaciju procesa donošenja odluka. Drvo odlučivanja može učinkovito modelirati složene nelinearne odnose i pružiti jasne uvide u to kako različite značajke utječu na izlazne vrijednosti.

Evaluacija regresijskih modela ključna je za razumijevanje njihove učinkovitosti. Uobičajene metrike uključuju srednju apsolutnu pogrešku (MAE) i srednju kvadratnu pogrešku (MSE), koje mjere prosječnu razliku između stvarnih i predviđenih vrijednosti. Koeficijent determinacije (R²) pokazuje koliko dobro prediktori objašnjavaju varijaciju izlazne varijable.

Regresija ima široku primjenu u raznim domenama. U ekonomiji, koristi se za predviđanje cijena dionica ili nekretnina. U medicini, može pomoći u predviđanju ishoda liječenja na temelju pacijentovih karakteristika. U sportu, koristi se za predviđanje rezultata utakmica na temelju statistika igrača. Također se primjenjuje u meteorologiji za prognozu vremenskih uvjeta i u mnogim drugim područjima gdje je potrebno kvantitativno predviđanje.

U zaključku, regresija kao kategorija nadziranog učenja pruža moćan alat za analizu i predviđanje kontinuiranih podataka. Njena sposobnost da iz velikih količina podataka izvuče vrijedne uvide čini je neizostavnim alatom u modernoj analizi podataka i donošenju informiranih odluka.

# ANALIZA PODATAKA

Podaci za strojno učenje čine temelj za treniranje, testiranje i evaluaciju modela strojnog učenja. Ovi podaci mogu biti strukturirani ili nestrukturirani. Strukturirani podaci su organizirani u tabličnom formatu, što omogućava njihovu jednostavnu obradu matematičkim funkcijama. Takvi podaci su često korišteni u strojnom učenju za zadatke klasifikacije i regresije. S druge strane, nestrukturirani podaci nemaju jasnu organizaciju, što otežava njihovu obradu i analizu. Strukturirani podaci obuhvaćaju, na primjer, Excel tablice ili SQL baze podataka, dok nestrukturirani podaci uključuju tekstualne dokumente, slike, videozapise i druge oblike podataka bez definirane strukture.

Za potrebu izrade projektnog zadatka korišten je Dataset “used\_device\_data” spremljen kao CSV datoteka s atributima:

Brand\_name, os, screen\_size, 4g, 5g, main\_camera\_mp, selfie\_camera\_mp, int\_memory, ram, battery, weight, release\_year, days\_uses, new\_price, used\_price

A black background with green and yellow text

Description automatically generated

Slika 3.1. Kod za učitavanje podataka

A black background with white text

Description automatically generated

Slika 3.2. Kod za prikaz prvih par redova podataka

A screenshot of a computer

Description automatically generated

Slika 3.3. Prikaz prvih 5 učitanih podataka

A screenshot of a computer

Description automatically generated

Slika 3.4. Kod i rezultat koji govori da dataset ima 3454 reda i 15 stupaca

A screenshot of a computer

Description automatically generated

Slika 3.5. Prikaz općih informacija

Postoje 3 tipa podataka u datasetu: integer, float i string.

A screenshot of a graph

Description automatically generated

Slika 3.6. Prikaz opisa podataka

Iz slike 3.6. možemo vidjeti raspon podataka i njhove srednje vrijednosti za svaki od atributa. Npr. veličina zaslona kreće se od 5 cm do 31 cm, s prosjekom od 13.7 cm, dok se rezolucija zadnje kamere kreće se od 0.08 do 48 MP, s prosječnom vrijednosti od 9.5 MP. Godine proizvodnje mobilnih uređaja su od 2013. do 2020. godine. U prosjeku su mobiteli rabljeni 675 dana, što je malo manje od dvije godine.

A screen shot of a computer

Description automatically generated

Slika 3.7. Provjera postoje li duplikati

Potrebno je odraditi provjeru nalazi li se u našim podatcima neka null vrijednost, odnosno nedostaje li neki od podataka, te prebrojati koliko postoji takvih podataka.

A screenshot of a computer

Description automatically generated

Slika 3.8. Prikaz koliko ima null podataka i pod kojim atributom se nalaze

## Eksplorativna analiza podataka

Eksplorativna analiza podataka (Exploratory Data Analysis, EDA) je ključan korak u analizi podataka koji pomaže razumjeti osnovne karakteristike skupa podataka. Kroz EDA, koriste se različite statističke metode i vizualizacije kako bi se identificirali obrasci, anomalije i relacije unutar podataka. Cilj je dobiti osjećaj za podatke prije nego što se pristupi složenijim analitičkim tehnikama. Uobičajeni alati uključuju grafikone poput histograma, dijagrama raspršenja i kutija dijagrama. EDA pomaže u otkrivanju osnovnih struktura, provjeri pretpostavki i donošenju odluka o daljnjoj analizi. Kroz ovaj proces, analitičari mogu bolje razumjeti podatke i donijeti informirane zaključke.

A screenshot of a computer

Description automatically generated

Slika 3.9. Proizvođači mobilnih uređaja

A screenshot of a graph

Description automatically generated

Slika 3.10. Proizvođači mobilnih uređaja u obliku grafa

U dataset je uključeno 34 proizvođača mobilnih uređaja, uključujući „Others“, koji podrazumijeva više manjih proizvođača. Ostali proizvođači zauzimaju 14.5% tržišta rabljenih mobilnih uređaja. U prvih 5 najzastupljenijih su svjetski poznate tvrtke kao Samsung, Huawei, LG i Lenovo. Apple, kao jedna od najpoznatijih tvrtki za mobilne uređaje, zauzima 1.1% tržišta rabljenih mobilnih uređaja, što sugerira da korisnici više vole kupovati nove uređaje.

A screenshot of a graph

Description automatically generated

Slika 3.11. Distribucija veličine ekrana po broju uređaja

A screenshot of a computer

Description automatically generated

Slika 3.12. Prikaz količine radne memorije po uređaju

Preko 80% uređaja ima 4GB radne memorije. Manje su zastupljene veličine od 6GB, 8GB i 2GB.

A graph of different colored bars

Description automatically generated

Slika 3.13. Prikaz godine proizvodnje po uređaju

A screen shot of a computer screen

Description automatically generated

Slika 3.14. Dijagram raspršenosti s parametrima težine i baterije

Iz Slike 3.14. pomoću dijagrama raspršenosti vidimo kako jačina baterije, utječe na povećanje težine mobilnog uređaja ovisno u marki uređaja.

# Priprema podataka

Priprema podataka ključni je korak u strojnom učenju. To uključuje čišćenje i transformaciju podataka kako bi bili konzistentni i korisni za model. Kvalitetna priprema podataka poboljšava performanse modela, osiguravajući preciznije predikcije i pouzdanije rezultate.

## Podatci koji nedostaju

Za početak, moramo provjeriti nedostaje li koji od podataka, odnosno je li nekom podatku vrijednost null.

A screenshot of a computer

Description automatically generated

Slika 4.1. Podatci koji imaju null vrijednosti

Kreiramo kopiju podataka kako bi izbjegli neželjene promjene.



Slika 4.2. Kod za kopiranje podataka u novu varijablu

A screen shot of a computer

Description automatically generated

Slika 4.3. Prvi korak rješavanja podataka koji nedostaju

Za varijable main\_camera\_mp, selfie\_camera\_mp, int\_memory, ram, battery i weight ubacit ćemo vrijednosti koje nedostaju u svakom stupcu s medijanom grupiranim prema brand\_name i release\_year budući da oba stupca imaju iskrivljenu distribuciju.

Vidimo da to nije rješilo problem jer i dalje postoje neki podatci koji nedostaju. Sada ćemo pokušati ubaciti vrijednosti koje nedostaju s medijanom grupiranim samo prema brand\_name.

A screenshot of a computer

Description automatically generated

Slika 4.4. Drugi korak rješavanja podataka koji nedostaju

Kako još ima podataka koji nedostaju u main\_camera\_mp, ubacit ćemo preostale vrijednosti koje nedostaju s medijanom za main\_camera\_mp.

A screenshot of a computer

Description automatically generated

Slika 4.5. Rješeni svi podatci koji nedostaju

Nakon zadnjeg koraka, sve vrijednosti koje nedostaju su rješene.

## Inženjering značajki

Inženjering značajki ("feature engineering") je proces stvaranja novih značajki iz sirovih podataka kako bi se poboljšala učinkovitost modela strojnog učenja. Kvalitetne značajke mogu značajno povećati točnost predikcija modela, omogućujući mu bolje razumijevanje i analizu podataka.

U našem slučaju, ima puno više smisla koristiti broj godina koji je prošao od proizvodnje mobilnog uređaja nego godinu proizvodnje. Stvoriti ćemo novu varijablu years\_since\_release tako što ćemo oduzeti release\_year od 2021. godine, to je godina u kojoj su podatci prikupljeni.

A screenshot of a computer

Description automatically generated

Slika 4.6. Prikaz podataka s dodanom novom varijablom

## Priprema podataka za modeliranje

Želimo predviditi normalized\_used\_price. Prije nego što nastavimo s izgradnjom modela, morat ćemo kodirati kategoričke značajke. Podijeliti ćemo podatke na trening podatke i testne podatke kako bismo mogli evaluirati model koji gradimo na testnim podatcima.

A screen shot of a computer screen

Description automatically generated

Slika 4.7. Definiranje x i y varijabli

A screenshot of a computer program

Description automatically generated

Slika 4.8. Podjela podataka na podatke za trening i podatke za testiranje

# Evaluacija rješenja

Evaluacija rješenja je od ključne važnosti za svaki projekt kako bi ocjenili uspješnost korištenih modela. Detaljna analiza pomaže identificirati pozitivne i negativne strane, omogućujući daljnje poboljšanje i prilagodbu rješenja prema ciljevima. Metrike kojima ćemo ocijenjivati performanse modela su:

* MSE (Mean Squared Error) - Srednja kvadratna pogreška, koja mjeri prosječnu kvadratnu razliku između stvarnih i predviđenih vrijednosti. Niže vrijednosti ukazuju na bolju preciznost.
* MAE (Mean Absolute Error) - Prosječna apsolutna pogreška, koja mjeri prosječnu apsolutnu razliku između stvarnih i predviđenih vrijednosti. Niže vrijednosti ukazuju na bolju preciznost.
* R2 - Koeficijent determinacije, koji pokazuje koliko dobro predviđene vrijednosti odgovaraju stvarnim vrijednostima. Vrijednost bliža 1 ukazuje na bolju prilagodbu modela.

## Linearna regresija

Linearna regresija je statistička metoda koja se koristi za predviđanje vrijednosti jedne varijable na osnovi vrijednosti druge varijable. Jednostavno rečeno, to je način da nacrtamo ravnu liniju kroz skup podataka tako da možemo vidjeti trend i predvidjeti buduće vrijednosti. Ova tehnika je vrlo korisna u različitim područjima, kao što su ekonomija, biologija i inženjering, jer nam pomaže razumjeti i kvantificirati odnos između varijabli.

A computer screen with text and numbers

Description automatically generated

Slika 5.1. Linearna regresija

Inicijaliziramo model linearne regresije koristeći klasu LinearRegression iz biblioteke scikit-learn. Zatim treniramo model (fit metoda) koristeći skup podataka za treniranje (x\_train i y\_train). Nakon treniranja, model koristimo za predviđanje vrijednosti na temelju skupa podataka za testiranje (x\_test).

Rezultati evaluacije za naš model linearne regresije su sljedeći:

* MSE: 23.575
* MAE: 16.676
* R2: 0.817

Ovi rezultati ukazuju na to da model linearne regresije dobro predviđa vrijednosti s obzirom na dostupne podatke. MSE i MAE su relativno niske, što znači da su predviđene vrijednosti bliske stvarnim vrijednostima. R2 od 0.817 pokazuje da model objašnjava 81.7% varijabilnosti u podacima, što ukazuje na visoku razinu prilagodbe modela.

## Random Forest Regressor

Random Forest Regressor je napredna metoda strojnog učenja koja koristi zbirku (šumu) odluka stabala za predviđanje vrijednosti ciljne varijable. Ova tehnika funkcionira tako da generira mnogobrojna stabla odluka tijekom treninga i koristi prosjek njihovih predviđanja za konačni rezultat. Random Forest je poznat po svojoj sposobnosti da rukuje složenim i nelinearnim odnosima između varijabli te je otporan na pretreniranje, što ga čini vrlo moćnim i pouzdanim alatom za regresiju u različitim domenama.

A computer screen with text and numbers

Description automatically generated

Slika 5.2. Random Forest Regressor

Inicijaliziramo model Random Forest Regressora koristeći klasu RandomForestRegressor iz biblioteke scikit-learn. Zatim treniramo model (fit metoda) koristeći skup podataka za treniranje (x\_train i y\_train). Nakon treniranja, model koristimo za predviđanje vrijednosti na temelju skupa podataka za testiranje (x\_test).

Rezultati evaluacije za naš Random Forest Regressor model su sljedeći:

* MSE: 24.655
* MAE: 18.877
* R2: 0.800

Ovi rezultati ukazuju na to da Random Forest Regressor model dobro predviđa vrijednosti s obzirom na dostupne podatke. MSE i MAE su relativno niske, što znači da su predviđene vrijednosti bliske stvarnim vrijednostima. R2 od 0.800 pokazuje da model objašnjava 80% varijabilnosti u podacima, što ukazuje na visoku razinu prilagodbe modela.

## Stablo odlučivanja

Stablo odlučivanja je metoda strojnog učenja koja koristi hijerarhijsku strukturu za donošenje predikcija. Ova tehnika razlaže podatke na manje podskupove kroz niz odluka temeljenih na značajkama podataka. Na svakom čvoru stabla donosi se odluka koja vodi do sljedećeg čvora ili završnog lista s predikcijom. Stabla odlučivanja su intuitivna i jednostavna za interpretaciju, te se često koriste u različitim domenama za klasifikaciju i regresiju zbog svoje jednostavnosti i jasnoće.

A computer screen shot of a program code

Description automatically generated

Slika 5.3. Stablo odlučivanja

Inicijaliziramo model Stabla Odlučivanja koristeći klasu DecisionTreeRegressor iz biblioteke scikit-learn. Zatim treniramo model (fit metoda) koristeći skup podataka za treniranje (x\_train i y\_train). Nakon treniranja, model koristimo za predviđanje vrijednosti na temelju skupa podataka za testiranje (x\_test).

Rezultati evaluacije za naš model Stabla Odlučivanja su sljedeći:

* MSE: 33.372
* MAE: 21.924
* R2: 0.634

Ovi rezultati ukazuju na to da model Stabla Odlučivanja dobro predviđa vrijednosti s obzirom na dostupne podatke. MSE i MAE su relativno visoke u usporedbi s prethodnim modelima, što ukazuje na manju preciznost. R2 od 0.634 pokazuje da model objašnjava 63.4% varijabilnosti u podacima, što je manje od prethodnih modela.

## K-najbližih susjeda

Metoda K Najbližih Susjeda je jednostavna, ali moćna tehnika nadziranog učenja koja se koristi za klasifikaciju i regresiju. Ova metoda se temelji na pretpostavci da slični primjeri imaju slične oznake ili vrijednosti. KNN algoritam klasificira ili predviđa vrijednost novog primjera na temelju oznaka ili vrijednosti njegovih najbližih susjeda u skupu podataka za učenje. Ovaj pristup je intuitivan i jednostavan za implementaciju, što ga čini popularnim u različitim područjima kao što su medicina, financije i prepoznavanje uzoraka.

A computer screen with white and black text

Description automatically generated

Slika 5.4. K-najbližih susjeda

Prvo, inicijaliziramo model metode K Najbližih Susjeda koristeći klasu KNeighborsRegressor iz biblioteke scikit-learn. Zatim treniramo model (fit metoda) koristeći skup podataka za treniranje (x\_train i y\_train). Nakon treniranja, model koristimo za predviđanje vrijednosti na temelju skupa podataka za testiranje (x\_test).

Rezultati evaluacije za naš model metode K Najbližih Susjeda su sljedeći:

* MSE: 28.373
* MAE: 20.384
* R2: 0.736

Ovi rezultati pokazuju da metoda k Najbližih Susjeda (KNN) dobro predviđa vrijednosti s obzirom na dostupne podatke. Srednja kvadratna pogreška (MSE) i prosječna apsolutna pogreška (MAE) su relativno niske, što ukazuje na bolju preciznost. R2 od 0.736 pokazuje da model objašnjava 73.6% varijabilnosti u podacima, što ukazuje na relativno visoku razinu prilagodbe modela.

# Zaključak

Procjena cijene rabljenih mobilnih uređaja zahtijeva pažljiv odabir algoritama strojnog učenja kako bi se postigla optimalna točnost i pouzdanost predikcija. U ovom istraživanju usporedili smo četiri različita modela: linearnu regresiju, stablo odlučivanja, Random Forest i metodu k najbližih susjeda (KNN).

Prije svega, potrebno je podatke, odnosno skup podataka, učitati, analizirati i dobro pripremiti za izgradnju modela. Korišteni su strukturirani podatci jer je riječ o regresiji. Eksplorativna analiza podataka je vrlo bitna kako bi dobili osjećaj za podatke prije pristupa složenijim analitičkim tehnikama. Od alata, od pomoći su bili grafikoni, histogrami te dijagrami raspršenja. Kod pripreme podataka bitno je provjeriti nedostaju li koji podatci, stvoriti nove značajke iz sirovih podataka ako za to ima potrebe te na kraju podijeliti podatke na testne i na trening podatke. Na kraju je potrebno evaluirati rješenja.

Linearna regresija pokazala se kao najprecizniji model za ovaj zadatak. Njena jednostavnost i sposobnost da linearno modelira odnose među značajkama rezultirala je visokom točnošću predikcija. Ovaj model je dobro prilagođen za slučajeve gdje postoji linearan odnos između značajki i ciljne varijable, što je često slučaj kod procjene cijena na temelju specifičnih karakteristika mobilnih uređaja.

Stablo odlučivanja nudi intuitivnu interpretaciju kroz hijerarhijsku strukturu odlučivanja, ali je pokazalo slabije performanse u usporedbi s drugim modelima. Glavni razlog tome je sklonost stabala odlučivanja prema pretreniravanju, gdje model previše prilagođava podatke za treniranje i gubi na generalizacijskoj sposobnosti za nove podatke. To ga čini manje pouzdanim za predikcije cijena rabljenih mobilnih uređaja.

Random Forest, koji koristi skupinu stabala odlučivanja, pokazao je značajno poboljšanje u odnosu na pojedinačno stablo odlučivanja. Ovaj model uspješno smanjuje problem pretreniravanja kombiniranjem više stabala i prosječenjem njihovih rezultata. Iako nije bio toliko precizan kao linearna regresija, Random Forest se pokazao kao vrlo robustan model za složenije odnose među značajkama i varijabilnost podataka.

Metoda k najbližih susjeda (KNN) oslanja se na usporedbu novih primjera s najbližim susjedima iz skupa podataka za treniranje. Iako KNN može biti vrlo efikasan u nekim scenarijima, njegova preciznost je bila niža u ovom slučaju. To može biti zbog osjetljivosti na izbor broja susjeda i utjecaja različitih razmjera značajki.

Zaključno, linearna regresija se pokazala kao najbolji model za procjenu cijena rabljenih mobilnih uređaja zahvaljujući svojoj točnosti i jednostavnosti. Random Forest također predstavlja pouzdan izbor zbog svoje otpornosti na pretreniravanje i sposobnosti rukovanja složenijim podatkovnim odnosima. Stablo odlučivanja i KNN su pokazali slabije performanse, te bi njihova primjena zahtijevala dodatna podešavanja i optimizacije kako bi se postigla bolja točnost predikcija.